

Αθήνα, 4 Μαΐου 2023

**Προς:** ΔΕΗ Α.Ε.  
Γενική Δ/νση Έργων & Εμπειρίας Πελάτη  
Δ/νση Στρατηγικής Δεδομένων & Ανάλυσης  
  
Υπόψη κ. Σταυριάννας Δημητρίου

**Θέμα:** Παροχή Υπηρεσιών Καθαρισμού & Επιβεβαίωσης Δεδομένων Πελατείας  
*Επικαιροποιημένη Πρόταση*

## 1. ΕΙΣΑΓΩΓΗ

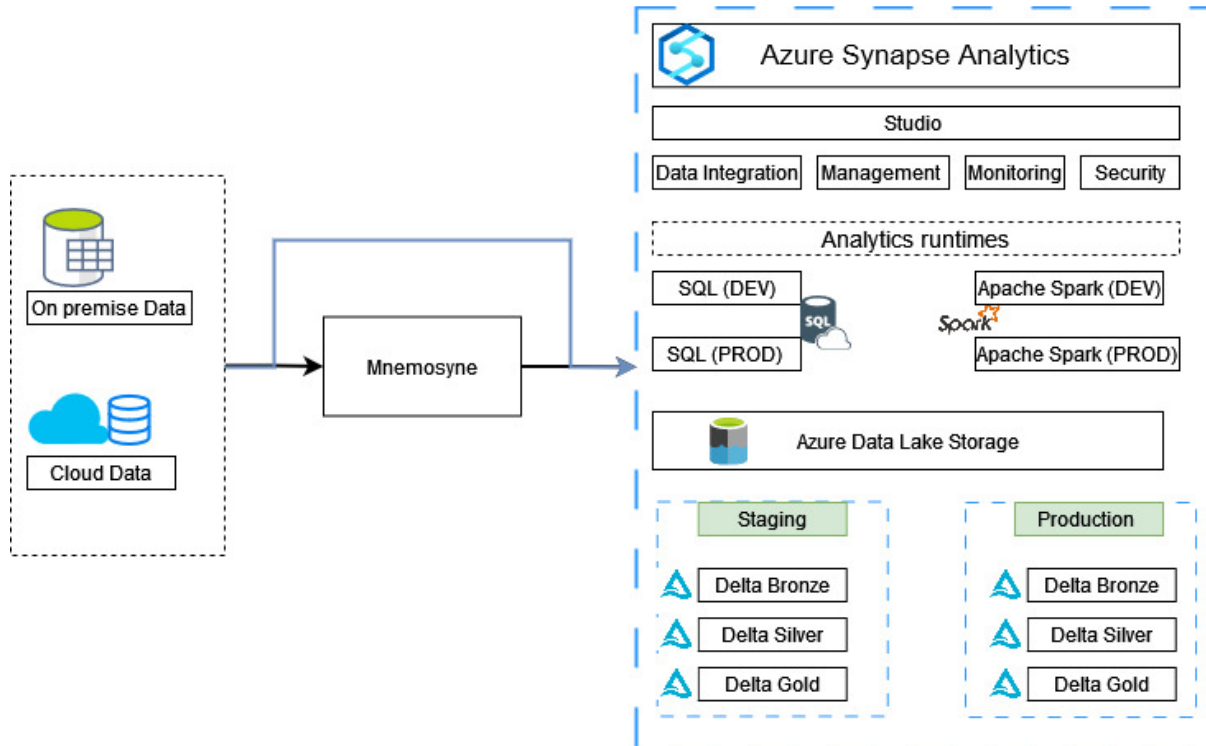
Η παρούσα προσφορά συντάσσεται στα πλαίσια αιτήματος της ΔΕΗ για παροχή υπηρεσιών καθαρισμού και επιβεβαίωσης στοιχείων πελατείας της.

Έχοντας υλοποιηθεί μέρος της διαδικασίας αυτής, διαπιστώθηκε ότι προκύπτουν συγκεκριμένες ανάγκες και επιπλέον εργασίες που κοινοποιήθηκαν στη Wemetrix ώστε να διαπιστωθεί σε ποιο βαθμό αυτή μπορεί να ανταποκριθεί στις απαιτούμενες υλοποιήσεις.

Η Wemetrix, διαθέτει τους κατάλληλους ανθρώπινους πόρους, την εμπειρία και την τεχνογνωσία που απαιτείται για υλοποίηση των ζητούμενων καθώς και για την έγκαιρη ολοκλήρωση των παραδοτέων του

## 2. ΔΟΜΗ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΕΡΓΑΣΙΑΣ

Σύμφωνα να την κατανόησή μας από τις περιγραφόμενες ανάγκες και εσωτερικές ροές της ΔΕΗ στο έντυπο προδιαγραφών του έργου, παραθέτουμε το flow chart της προσέγγισης μας



Το Azure Synapse Analytics είναι μια ισχυρή και ενοποιημένη υπηρεσία ανάλυσης δεδομένων που παρέχεται στο Microsoft Azure. Συνδυάζει την επεξεργασία μεγάλων δεδομένων και τις δυνατότητες αποθήκευσης δεδομένων σε ένα μόνο πακέτο λύσης, επιτρέποντας στις οργανώσεις να αναλύουν μεγάλους όγκους δεδομένων αποδοτικά και να αποκτήσουν πολύτιμες εισηγήσεις. Η αρχιτεκτονική της ανάλυσης δεδομένων στο Azure Synapse περιλαμβάνει αρκετά βασικά στοιχεία που συνεργάζονται για την παροχή ολοκληρωμένων δυνατοτήτων ανάλυσης.

- **Εισαγωγή Δεδομένων:** Η διαδικασία ανάλυσης δεδομένων ξεκινά με την εισαγωγή δεδομένων από διάφορες πηγές στο Azure Synapse. Αυτό μπορεί να περιλαμβάνει δομημένα, ημιδομημένα και αδομητά δεδομένα από βάσεις δεδομένων, πλατφόρμες και άλλες εξωτερικές πηγές. Το Azure Synapse υποστηρίζει μια ευρεία γκάμα μεθόδων εισαγωγής, όπως το Azure Data Factory, η αποθήκευση του Azure Data Lake, η αποθήκευση του Azure Blob κ.λπ. Τα δεδομένα μεταφέρονται στον χώρο εργασίας Synapse, όπου μπορούν να προετοιμαστούν για περαιτέρω επεξεργασία.

- **Προετοιμασία και Μετασχηματισμός Δεδομένων:** Μόλις τα δεδομένα εισαχθούν στο Azure Synapse, υποβάλλονται σε βήματα προετοιμασίας και μετασχηματισμού για να γίνουν κατάλληλα για ανάλυση. Το Azure Synapse παρέχει ισχυρές δυνατότητες ενσωμάτωσης δεδομένων μέσω εργαλείων όπως το Azure Data Factory, το Azure Databricks και οι SQL Server Integration Services (SSIS). Αυτά τα εργαλεία επιτρέπουν τον καθαρισμό των δεδομένων, την επεξεργασία τους και τον μετασχηματισμό τους, επιτρέποντας στις οργανώσεις να διαμορφώσουν τα δεδομένα σύμφωνα με τις αναλυτικές απαιτήσεις τους.
- **Αποθήκευση και Διαχείριση Δεδομένων:** Το Azure Synapse αξιοποιεί δύο βασικά στοιχεία για την αποθήκευση και τη διαχείριση δεδομένων: το Azure Data Lake Storage και οι κατανεμημένες αποθηκευτικές δυνατότητες βασισμένες σε SQL. Το Azure Data Lake Storage λειτουργεί ως ένα εξαιρετικά κλιμακούμενο και οικονομικό αποθηκευτικό χώρο για την αποθήκευση μεγάλων όγκων ακατέργαστων και επεξεργασμένων δεδομένων. Παρέχει έναν ιεραρχικό αρχαιοθέτη που μπορεί να χειριστεί διάφορους τύπους και μορφές δεδομένων. Από την άλλη πλευρά, οι κατανεμημένες αποθήκευσης δεδομένων βασισμένες σε SQL, γνωστές ως Azure Synapse SQL Pools, επιτρέπουν τη δημιουργία και διαχείριση αποθήκων δεδομένων που είναι βελτιστοποιημένες για αναλυτικές εργασίες. Παρέχει δυνατότητες ερωτήσεων υψηλής απόδοσης και υποστηρίζει τυπικές ερωτήσεις SQL.
- **Επεξεργασία Δεδομένων και Ανάλυση:** Το Azure Synapse προσφέρει πολλές επιλογές για την επεξεργασία δεδομένων και την ανάλυση, καλύπτοντας διάφορες χρήσεις και απαιτήσεις φορτίου εργασίας. Ένα από τα βασικά στοιχεία είναι το Apache Spark, που παρέχει δυνατότητες κατανεμημένης επεξεργασίας για εργασίες μεγάλων δεδομένων. Το Azure Synapse χρησιμοποιεί το Azure Databricks, μια πλατφόρμα ανάλυσης βασισμένη στο Apache Spark, για να επιτρέψει την κλιμακούμενη εξερεύνηση δεδομένων, την προετοιμασία δεδομένων και την προηγμένη ανάλυση. Επιπλέον, οι κατανεμημένες αποθήκες δεδομένων Azure Synapse SQL παρέχουν ένα βελτιστοποιημένο περιβάλλον για την εκτέλεση παραδοσιακών ερωτημάτων SQL σε δομημένα και ημιδομημένα δεδομένα.
- **Οργάνωση και Παρακολούθηση:** Το Azure Synapse ενσωματώνεται με το Azure Data Factory για να παρέχει δυνατότητες οργάνωσης δεδομένων. Το Azure Data Factory επιτρέπει τη δημιουργία ροών για τον προγραμματισμό και οργάνωση της εκτέλεσης διαφόρων εργασιών ενσωμάτωσης και μετασχηματισμού δεδομένων. Παρέχει ένα οπτικό περιβάλλον για τον σχεδιασμό, την παρακολούθηση και τη διαχείριση των αγωγών δεδομένων,

εξασφαλίζοντας την ομαλή ροή των δεδομένων κατά τη διάρκεια της αναλυτικής διαδικασίας. Το Azure Synapse περιλαμβάνει επίσης χαρακτηριστικά παρακολούθησης και διαχείρισης που επιτρέπουν στις οργανώσεις να παρακολουθούν την απόδοση, την κλιμάκωση και την υγεία των διαδικασιών ανάλυσης δεδομένων.

Η προτεινόμενη αρχιτεκτονική της ανάλυσης δεδομένων στο Azure Synapse παρέχει ένα ισχυρό περιβάλλον για την ανάπτυξη και εκτέλεση αναλυτικών εργασιών σε μεγάλη κλίμακα. Η τελική αρχιτεκτονική της λύσης θα καταγραφεί με περισσότερη λεπτομέρεια κατά τη διάρκεια της πρώτης εβδομάδας και αφού πρώτα έχουμε την πλήρη εικόνα της υφιστάμενης λύσης/υποδομής.

Η πρόταση/πρόθεση μας είναι να διατηρηθεί η υπάρχουσα λύση και όλες οι διεργασίες που γίνονται σε αυτή μέχρι το σημείο της ομαδοποίησης του golden record id. Το στάδιο αυτό (golden record ID ) θα υλοποιηθεί από την αρχή με την προτεινόμενη λύση και χρήση του Delta Lake μηχανισμού.

Όλα τα παραπάνω θα αποσαφηνιστούν και θα συν-αποφασιστούν μετά την ολοκλήρωση των workshops όπου θα υπάρχει πλήρης κατανόηση της υπάρχουσας υποδομής ( θα συν – υπολογιστεί και ο χρόνος υλοποίησης σε κάθε μια από τις 2 αρχιτεκτονικές, existing & suggested )

### 3. ΠΡΟΤΑΣΗ ΥΛΟΠΟΙΗΣΗΣ / TASKS

Στον παρακάτω πίνακα αποτυπώνονται οι απαιτούμενες εργασίες και οι προδιαγραφές που έχουν ζητηθεί καθώς και η μεθοδολογική προσέγγισή μας για τις υλοποιήσεις στα πλαίσια του έργου.

Για την υλοποίηση των παρακάτω θα χρησιμοποιηθεί Apache Spark (PySpark, Spark SQL) μέσω του Synapse.

Εργασία	Υπό εργασία	Περιγραφή Εργασίας	Μεθοδολογία
1	1.1	Αξιολόγηση IIS & Migration σε Synapse	Θα εξεταστούν ενδελεχώς οι διαδικασίες που εκτελούνται αυτή τη στιγμή από την υπάρχουσα υποδομή. Σε συνεργασία με την ομάδα της ΔΕΗ θα συν αποφασιστεί για το ποιες διαδικασίες μπορούν να μεταφερθούν στην νέα αρχιτεκτονική και να καταργηθούν από την παλιά. Από την διαδικασία αυτή εξαιρείται το grouping του Golden Record ID , το οποίο θα υλοποιηθεί έτσι κι αλλιώς με την προτεινόμενη λύση

			<p>με σκοπό και την αξιοποίηση του Delta Lake και την σταθεροποίηση του golden record ID.</p> <p>Υ.Γ : Ανάλογα με το ποιες διαδικασίες θα μπουν στο migration , ενδέχεται να επηρεαστεί και το πλάνο υλοποίησης. Θα αποσαφηνιστεί κατά την διάρκεια της 1<sup>ης</sup> εβδομάδας , όπου θα έχει γίνει και παρουσίαση της υπάρχουσας υποδομής από την ομάδα της ΔΕΗ</p>
2	2.1	Λεπτομερές χρονοδιάγραμμα και παραδοτέα	Ανάλογα με τις διαδικασίες migration που ενδεχομένως προκύψουν , θα δοθεί λεπτομερές χρονοδιάγραμμα εκτέλεσης διεργασιών με βάση και του προγραμματισμού του CRM migration που ακολουθεί
3	3.1	Προσθήκη Φύλου Αντί-Συμβαλλόμενου (Gender)	<p>Αντιστοίχιση του ονόματος με βάση δεδομένων που περιέχει όνομα και φύλλο και δυνατότητα διόρθωσης του ονόματος σε περίπτωση λανθασμένης καταχώρησης με την εύρεση του πιο συναφούς ονόματος από τη βάση.</p> <p>Ενδεικτικά :</p> <ul style="list-style-type: none"> <li>Αξιοποίηση ανοιχτών βάσεων δεδομένων (UCI: University of California Irvine) για λατινικούς χαρακτήρες</li> </ul>
4	4.1	Αναγνώριση και διόρθωση email	Αναγνώριση διεύθυνσης email πελάτη, εντοπισμός παρόχου email και διόρθωση τυπογραφικών με χρήση regular expressions
5	5.1	Αναγνώριση και διόρθωση ΑΤ (στρατιωτικής, αστυνομικής, πολιτικής) & Διαβατήριο (Ελληνικό ή Ξένο)	Regular expressions, έλεγχος format, έλεγχος αριθμού ψηφίων / λεκτικών (length validation)
6	6.1	Αναγνώριση είδους πελάτη (customer type):	<p>Εφαρμογή κανόνων με βάση το VAT.</p> <p>Εναλλακτικές προτάσεις θα προκύψουν μετά την ολοκλήρωση των workshops.</p>
7	7.1	Αναγνώριση ορθότητας ευρέως χρησιμοποιούμενων ΑΦΜ με βάση τα λεκτικά	<p>Δημιουργία dimension πίνακα στον οποίο θα διατηρούνται τα ευρέως χρησιμοποιούμενα ΑΦΜ.</p> <p>Έλεγχος κάθε ΑΦΜ με τον πίνακα ( join , case when expressions) και επαλήθευση ή ακύρωση (αντικατάσταση με «null») του πεδίου</p>
8	8.1	Αναγνώριση και απομόνωση εγγραφών CAs με συγκεκριμένα λεκτικά	Αντιστοίχιση με δεδομένη λίστα εξαιρέσεων

9	9.1	Αντικατάσταση κενών Billing Addresses με POD Addresses	Εφαρμογή διαδικασιών Golden Records στα PoD addresses και συμπλήρωση των billing με βάση των PoD
10	10.1	Κανονικοποίηση ονομάτων B2B_Company_Name	Επιλογή «κανονικών» ονομάτων και καταχώρηση σε λίστα. Αντιστοίχιση με το πιο συναφές company name με χρήση regular expressions
11	11.1	Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ για Ιδιώτες	Ορισμός ως «πολλαπλό» αν έχει αριθμό πολλαπλού και μετέπειτα ομαδοποίηση ή έλεγχος για πολλαπλούς λογαριασμούς με ίδιο ΑΦΜ στην περίπτωση κενού αριθμού πολλαπλού και ομαδοποίηση κατά κύριο ΑΦΜ.
	11.2	Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ για Δημόσιο	Παρόμοια διαδικασία με 9.1 συν αντιστοιχία με τη σημαία ένδειξης δημόσιο ή όχι από το δεδομένο πίνακα
	11.3.1	Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ για Δήμους – Καλλικράτης	Συσχέτιση με βάση το excel που περιέχει τους δήμους Καλλικράτη
	11.3.2	Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ για Δήμους – ΟΧΙ Καλλικράτης	Αντίστοιχο με 9.3.1 για εγγραφές που παραπέμπουν δε Δήμο αλλά δεν είναι στο αρχείο Καλλικράτη
	11.4	Golden Record Groups τύπου Πολλαπλοί «Πολλαπλοί»	Εφαρμογή regular expressions ή/και fuzzy matching τεχνικών στα λεκτικά των επωνυμιών για ομαδοποίηση πολλαπλών που αντιστοιχούν στον ίδιο πελάτη.  Για την υλοποίηση fuzzy matching (ενδεικτικά) με την χρησιμοποίηση existing βιβλιοθηκών: <ul style="list-style-type: none"> <li>• Levenshtein Distance</li> <li>• Similarity Score</li> <li>• Στατιστικά έγκυρων πεδίων (πόσα πεδία αξιοποιούνται για την κατηγοριοποίηση)</li> </ul>
12	12.1	Αναγνώριση περιπτώσεων Πολλαπλών με «λάθος καταχωρημένο ΑΦΜ»	Αντιστοίχιση βάση ονόματος και κατά συνέπεια ΑΦΜ ή αναζήτηση ΑΦΜ με απόκλιση μερικών (ενός-δύο) χαρακτήρων για την περίπτωση του ορθογραφικού.
13	13.1	Δημιουργία πεδίου «Μέθοδος Ομαδοποίησης» ειδικά για την περίπτωση Πολλαπλών ώστε να διαφοροποιείται η μέθοδος ομαδοποίησης	Δημιουργία τριών ξεχωριστών ομάδων όπως εκείνες περιγράφονται: <ul style="list-style-type: none"> <li>• «Ενταγμένος» με βάση ΑΦΜ ή Αρ_Πολλαπλού</li> <li>• «Ανένταχτος» με βάση «βασικό» ΑΦΜ Πολλαπλού</li> <li>• «Αλγόριθμος» με fuzzy matching τεχνικές</li> </ul>

		των CAs σε κάθε Golden Record Group	
14	14.1	Ένταξη στην διαδικασία matching για Golden Record Group Analysis πεδίων σχετικών με Ebill	Εφαρμογή κανόνων Golden Record για τα πεδία e-bill username και email
15	15.1	Ένταξη πεδίων σχετικών με Ταυτότητες, Διαβατήρια & E-bill στα Golden Record Group – level πεδία	Εφαρμογή κανόνων Golden Record για τα πεδία ταυτότητας, διαβατηρίου και e-bill username και email
16	16.1	Δημιουργία κοινής μεθοδολογίας συμπλήρωσης πεδίων επιπέδου Golden Record Group από τα υφιστάμενα πεδία επιπέδου CA	Σε περίπτωση διπλών εγγραφών, διατήρηση της πιο πρόσφατης με βάση την ημερομηνία έκδοσης λογαριασμού
17	17.1	Δημιουργία GR_Customer_Type που θα αφορά το είδος πελάτη του Golden Record Group και όχι του CA	Ομαδοποίηση πελατών βάση συγκεκριμένων πεδίων εκτός CA.
18	18.1	Υλοποίηση Extra post-processing requirements	Εφαρμογή business κανόνων που θα δοθούν στην ομάδα υλοποίησης (python / sql /spark)
19	19.1	Υποστήριξη για τις ανάγκες μεταφόρτωσης των δεδομένων στην νέα CRM πλατφόρμα	Ad-hoc support
20	20.1	Παροχή score ομοιότητας εγγραφών CAs που έχουν ομαδοποιηθεί στο ίδιο Golden Record Group σε κοινή κλίμακα αξιολόγησης	<p>Ανάπτυξη και υλοποίηση μεθοδολογίας ασαφής αντιστοίχισης κειμένου με σκοπό την τεχνική εύρεση συμβολοσειρών που ταιριάζουν με ένα μοτίβο κατά προσέγγιση και όχι ακριβώς (fuzzy matching) με την χρησιμοποίηση existing βιβλιοθηκών.</p> <p>Ενδεικτικά:</p> <ul style="list-style-type: none"> <li>• Levenshtein Distance</li> <li>• Similarity Score</li> </ul>

			<ul style="list-style-type: none"> <li>Στατιστικά έγκυρων πεδίων ( πόσα πεδία αξιοποιούνται για την κατηγοριοποίηση )</li> </ul>
21	21.1	Βελτίωση Ομαδοποίησης CAs και με βάση τεχνικές διαδικασίες fuzzy matching λεκτικών πεδίων CAs	Παρόμοια με 18.1 ( fuzzy matching) , scoring και εφαρμογή ορίων ( κατώφλι απόφασης ). Επιπλέον εφαρμογή regular expressions , όπου είναι εφικτό χρησιμοποιώντας λέξεις κλειδιά .
22	22.1	Δημιουργία αυτόματου μηχανισμού delta επικαιροποίησης	Δημιουργία Delta Lake σε Azure : <ul style="list-style-type: none"> <li>Διατήρηση ιστορικών εκδοχών των διαδοχικών runs</li> <li>Άμεσα προσβάσιμα δεδομένα</li> <li>Σύγκριση και δυνατότητα ανάλυσης αλλαγών και κατά συνέπεια δημιουργία versioning των Golden Record Groups</li> </ul>
23	23.1	Διατήρηση σταθερού Golden Record ID μεταξύ διαφορετικών runs του μηχανισμού τακτικής (εβδομαδιαίας) ανανέωσης	Διατήρηση μέσω διαφορετικών εκδοχών ( delta lake ) και/ ή δημιουργία dimension πίνακα στον οποίο θα διατηρούνται τα Golden Record IDs από κάθε run και με συνδυασμό ( join) θα εξασφαλίζουμε τη διατήρηση του κλειδιού.
24	24.1	Logging εφαρμογής	Δημιουργία logs από το περιβάλλον Azure. Ενδεικτικά : <ul style="list-style-type: none"> <li>Active directory logs</li> <li>Storage logs</li> <li>Activity logs</li> </ul> Παραμετροποίηση συχνότητας εγγραφής , διατήρηση ιστορικού και χώρου αποθήκευσης με βάση τη διαθεσιμότητα του περιβάλλοντος εργασίας.
	24.2	Logging διεργασιών	Προτείνουμε σαν extension του «κλασικού» logging της εφαρμογής, την καταγραφή και ανάλυση των διεργασιών cleansing ( υπολογισμοί , διορθώσεις, κατηγοριοποιήσεις ) μέσω περιγραφικών στατιστικών. (π.χ. ποσοστό διόρθωσης ΑΦΜ, ποσοστό πελατών που ανήκουν σε κάποια συγκεκριμένη κατηγορία) .

#### 4. ΟΜΑΔΑ ΕΡΓΑΣΙΑΣ



Η Wemetrix θα παρέχει ομάδα έμπειρων μηχανικών με σκοπό την υλοποίηση των διεργασιών όπως αναφέρονται στο διάγραμμα που προηγείται στα πλαίσια της χρόνου που απαιτείται για την ολοκλήρωση του έργου και το οποίο θα αποτυπωθεί λεπτομερώς στο χρονοδιάγραμμα (1<sup>ο</sup> παραδοτέο του έργου στο τέλος της πρώτης εβδομάδας). Η ομάδα θα είναι πλήρως αφοσιωμένη στην εκτέλεση του έργου. Πιο συγκεκριμένα η Wemetrix θα παρέχει:

- 3 senior Data Engineers
- 1 Data architect
- 1 Project / Delivery Manager

Όλη η ομάδα θα έχει πρόσβαση στο MS Azure cloud της ΔΕΗ. Όλα τα developments θα πραγματοποιούνται στο staging περιβάλλον και θα περνούν στο production μόνο μετά από έγκριση από την αρμόδια ομάδα της ΔΕΗ. “Way of working” θα συζητηθεί και θα συμφωνηθεί μεταξύ των ομάδων κατά τη πραγματοποίηση των workshops που θα πραγματοποιηθούν κατά τη διάρκεια της 1<sup>ης</sup> εβδομάδας.

## 5. ΠΑΡΑΔΟΤΕΑ

Η Wemetrix αναλαμβάνει να παραδώσει τα όσα αναφέρονται στο κείμενο των προδιαγραφών του έργου εντός του χρονικού πλαισίου υλοποίησης, και πιο συγκεκριμένα:

1. Λεπτομερές χρονοδιάγραμμα του έργου. Το ακριβές χρονοδιάγραμμα θα προκύψει μετά από τα αρχικά workshops και θα κατατεθεί επίσημα εντός της 1<sup>ης</sup> εβδομάδας του έργου. Για διευκόλυνση της αξιολόγησης της πρότασής μας, ενδεικτικό χρονοδιάγραμμα παρατίθεται στην Ενότητα 6 της παρούσας.
2. Τελικό consolidated output dataset με τα πιο πρόσφατα επικαιροποιημένα στοιχεία της βάσης δεδομένων της ΔΕΗ.
3. Ροές που υλοποιούν τις προδιαγραφές φορτωμένες και στο staging και production περιβάλλον στο MS Azure της ΔΕΗ σύμφωνα με τις διαδικασίες που ορίζονται από το έργο.
4. Μεταφορά στο staging περιβάλλον της ΔΕΗ όλων των τροποποιημένων υλοποιήσεων προς έλεγχο και UAT από την αρμόδια ομάδα της ΔΕΗ. Οριστικοποίηση και μεταφορά των αλλαγών σε περιβάλλον παραγωγής θα πραγματοποιείται μόνο μετά την επιβεβαίωση της ορθότητας και των ελέγχων UAT από την ομάδα της ΔΕΗ.
5. Έκθεση περιγραφής και τεκμηρίωσης της μεθοδολογίας και του κώδικα/ρών που θα παραδοθούν
6. Ενημέρωση και εκπαίδευση των χρηστών στελεχών της ΔΕΗ για τις τροποποιήσεις και τον τρόπο υλοποίησης.

## 6. ΕΝΔΕΙΚΤΙΚΟ ΧΡΟΝΟΔΙΑΓΡΑΜΜΑ

	Weeks				
	M0				M1
	W0	W1	W2	W3	W4
Εργασία 0: Παράδοση business κανόνων και πραγματοποίηση workshops					
Εργασία 0: Παρουσίαση της υπάρχουσας αρχιτεκτονικής από την ομάδα της ΔΕΗ					
Εργασία 1: Migration διαδικασιών εφόσον χρειάζεται - IIS evaluation					
Εργασία 2 :Λεπτομερές χρονο-διάγραμμα του έργου					
Εργασία 2 : Σχεδιασμός Διαδικασίας Ανάλυσης Δεδομένων					
Εργασία 3: Προσθήκη Φύλου Αντί-Συμβαλλόμενου (Gender)					
Εργασία 4: Αναγνώριση και διόρθωση email					
Εργασία 5: Αναγνώριση και διόρθωση AT (στρατιωτικής, αστυνομικής, πολιτικής) & Διαβατήριο (Ελληνικό ή Ξένο)					
Εργασία 6: Αναγνώριση είδους πελάτη (customer type):					
Εργασία 7: Αναγνώριση ορθότητας ευρέως χρησιμοποιούμενων ΑΦΜ με βάση τα λεκτικά					
Εργασία 8: Αναγνώριση και απομόνωση εγγραφών CAs με συγκεκριμένα λεκτικά					
Εργασία 9: Αντικατάσταση κενών Billing Addresses με POD Addresses					
Εργασία 10: Κανονικοποίηση ονομάτων B2B_Company_Name					
Εργασία 11: Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ					
Εργασία 12: Αναγνώριση περιπτώσεων Πολλαπλών με «λάθος καταχωρημένο ΑΦΜ»					
Εργασία 13: Δημιουργία πεδίου «Μέθοδος Ομαδοποίησης» ειδικά για την περίπτωση Πολλαπλών ώστε να διαφοροποιείται η μέθοδος ομαδοποίησης των CAs σε κάθε Golden Record Group					
Εργασία 14: Ένταξη στην διαδικασία matching για Golden Record Group Analysis πεδίων σχετικών με Ebill					
Εργασία 15: Ένταξη πεδίων σχετικών με Ταυτότητες, Διαβατήρια & E-bill στα Golden Record Group – level πεδία					
Εργασία 16: Δημιουργία κοινής μεθοδολογίας συμπλήρωσης πεδίων επιπέδου Golden Record Group από τα υφιστάμενα πεδία επιπέδου CA					
Εργασία 17: Δημιουργία GR_Customer_Type που θα αφορά το είδος πελάτη του Golden Record Group και όχι του CA					
Εργασία 18: Υλοποίηση Extra post-processing requirements					
Εργασία 19: Υποστήριξη για τις ανάγκες μεταφόρτωσης των δεδομένων στην νέα CRM πλατφόρμα	AD HOC REQUEST				
Εργασία 20: Παροχή score ομοιότητας εγγραφών CAs που έχουν ομαδοποιηθεί στο ίδιο Golden Record Group σε κοινή κλίμακα αξιολόγησης					
Εργασία 21: Βελτίωση Ομαδοποίησης CAs και με βάση τεχνικές διαδικασίες fuzzy matching λεκτικών πεδίων CAs					
Εργασία 22: Δημιουργία αυτόματου μηχανισμού delta επικαιροποίησης					
Εργασία 23: Διατήρηση σταθερού Golden Record ID μεταξύ διαφορετικών runs του μηχανισμού τακτικής (εβδομαδιαίας) ανανέωσης					
Εργασία 24: Logging System / Cleansing					
Εργασία 25: Admin / Documentation					

## 7. ΟΙΚΟΝΟΜΙΚΗ ΠΡΟΣΦΟΡΑ

Πάγια πολιτική καθορισμού των αμοιβών της Wemetrix για τις παρεχόμενες υπηρεσίες της και την ολοκλήρωση των παραδοτέων είναι το κόστος των αδειών χρήσης των τεχνολογικών λύσεων τις οποίες χρησιμοποιεί για λογαριασμό των πελατών της, ο χρόνος απασχόλησης όλων των βαθμίδων εμπειρίας των στελεχών της και η πολυπλοκότητα του έργου που αναλαμβάνει.

Για τους σκοπούς της παρούσας προσφοράς, η αμοιβή μας θα ανέλθει στα € 248.000 (διακόσιες σαράντα οκτώ χιλιάδες ευρώ) πλέον του αναλογούντος ΦΠΑ.

Η καταβολή του ποσού προτείνεται να πραγματοποιηθεί ως εξής:

30%, ήτοι ποσό € 74.400 (εβδομήντα τέσσερις χιλιάδες τετρακόσια ευρώ) πλέον ΦΠΑ με την αποδοχή της προσφοράς μας και την ανάθεση της υλοποίησης

70%, ήτοι ποσό € 173.600 (εκατό εβδομήντα τρεις χιλιάδες εξακόσια ευρώ) πλέον ΦΠΑ με την ολοκλήρωση όλων των παραδοτέων και την αποδοχή τους από τα αρμόδια στελέχη της ΔΕΗ.